## EXECUTIVE SUMMARY

General Introduction:

In 1998, the U.S. Environmental Protection Agency (EPA) requested that the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) evaluate the validation status of the Frog Embryo Teratogenesis Assay—*Xenopus* (FETAX). ICCVAM agreed to coordinate a review of the method, and the National Toxicology Program (NTP) Interagency Center for the Validation of Alternative Toxicological Methods (NICEATM) agreed to prepare a Background Review Document (BRD) summarizing the available data and the extent to which each of the ICCVAM validation and acceptance criteria have been met. NICEATM assessed the validation status of FETAX as a screening assay for detecting potential human teratogens, and for its use in the ecotoxicological assessment of water/soil/sediment samples.

Historical Background: Dr. James Dumont introduced FETAX, which uses the embryos of the South African clawed frog (Xenopus laevis), in 1983. Since its introduction, a number of inter-laboratory studies, largely directed by Drs. John Bantle (Oklahoma State University, Stillwater, OK, U.S.) and Douglas Fort (Stover Biometrics Laboratory, Stillwater, OK, U.S.), have been conducted to validate the utility of the assay for developmental hazard assessment. These validation studies were conducted in collaboration with the U.S. Army and the National Institute of Environmental Health Sciences (NIEHS). In 1991, the American Society for Testing and Materials (ASTM) developed a test guideline for FETAX, which was subsequently revised and republished in 1998. With regard to human developmental hazard assessment, this document reviews the information provided by 276 studies involving 137 substances. For ecotoxicological hazard assessment, test data from ten publications involving 124 water/soil/sediment samples were considered.

**FETAX for Human Developmental Hazard Identification**

Rationale: FETAX is a 96-hour *in vitro* whole-embryo test developed to determine the teratogenic and developmental toxicity potential of chemicals and complex mixtures. The primary endpoints include mortality, malformations, and growth inhibition. Based on mortality and malformation data obtained over a range of dose levels, the 50% lethal concentration ($LC_{50}$) (i.e., the concentration estimated to induce lethality in 50% of exposed embryos) and the 50% effective concentration for malformations ($EC_{50}$) (i.e., the concentration estimated to induce malformations in 50% of exposed embryos) are calculated. These two point estimates are used to calculate the teratogenic index (TI), which is equal to the $LC_{50}$ divided by the $EC_{50}$. Growth is ascertained by measuring the head to tail length of the embryos. The minimum concentration to inhibit growth (MCIG) (i.e., the lowest effective concentration for growth inhibition) is determined by statistically comparing the mean 96-hour head to tail length of the treated embryos at each treatment concentration to that of the control embryos. The statistical comparison is based on using student's t-test for grouped observations at the p=0.05 level. Any one of three criteria (TI, growth inhibition, or severity of induced malformations) is used to identify a teratogen. TI values greater than 1.5 signify a greater potential for embryos to be malformed in the absence of significant embryo mortality. Growth inhibition is stated to be correlated with teratogenicity in FETAX, and teratogenic hazard is considered to be present when growth is significantly inhibited at concentrations below 30% of the 96-hour $LC_{50}$ (i.e., when the MCIG/$LC_{50}$ ratio is less than 0.30). Teratogens generally cause moderate to severe malformations at concentrations near the 96-hour $LC_{50}$.

*Mechanistic Basis:* FETAX is essentially an organogenesis test, and organogenesis is highly conserved across amphibians and mammals. The first 96 hours of embryonic development in *Xenopus* parallel many of the major processes of human organogenesis. Thus, FETAX should be useful in predicting potential human developmental toxicants and teratogens. Due to the nature of the endpoints assessed, FETAX does not provide information on substances that may induce functional developmental deficits in mammals. As *Xenopus* embryos are deficient in mixed function oxidase-dependent metabolic activation processes, the addition of an exogenous

metabolic activation system (MAS) to the assay allows for an assessment of the developmental toxicity of metabolites in addition to the parent substance.

*Regulatory Rationale:* Current Federal regulations require determination of the developmental toxicity potential of many chemicals and products. EPA regulations specify the use of at least one, but usually two mammalian species (e.g., rats, mice, rabbits, hamsters) for the testing of fuels and fuel additives, pesticides, and other materials. The Organization for Economic Cooperation and Development (OECD) guidelines do not explicitly restrict developmental toxicity testing to mammals, although the use of FETAX has not been specifically addressed. Because FETAX is relatively easy, rapid, and inexpensive, the test has been proposed as a screening assay to identify potential human teratogens and developmental toxicants. As a screening test, a positive FETAX response would indicate a potential human hazard and, in the absence of other data, would be considered a presumptive teratogen/developmental toxicant. A negative FETAX response would not necessarily indicate the absence of a hazard, and negative responses would be followed by definitive *in vivo* mammalian testing A positive response would require no further testing unless there is concern about a potential false positive response (e.g., the positive FETAX response occurs at doses not applicable to the *in vivo* situation). For public agencies, such information could also be used to prioritize chemicals for more definitive testing. Regardless of the result obtained, an investigator may conclude that confirmatory testing is merited based on consideration of supplemental information, such as structure-activity relationships (SAR) and other chemical and/or testing data.

FETAX is considered to be applicable to all chemicals individually or in formulations, and to commercial products or mixtures that can be measured accurately at the necessary concentrations in water. This assay has not yet been considered for acceptance by U.S. Federal agencies for human health hazard assessment. The most commonly used protocol for identifying a potential human developmental hazard involves the administration of a test substance at three dose levels to pregnant laboratory mammals (usually mice, rats, or rabbits) during the period of major organogenesis. Treatment is followed by evaluation of maternal responses throughout pregnancy, and then examination of the dam and the uterine contents just prior to term. The developmental toxicity endpoints assessed include mortality (e.g., incidence of total, early, and

late fetal deaths), malformations (external, visceral, skeletal), variations (external, visceral, skeletal), growth (body weight), clinical signs (type, incidence, duration, and degree), and gross necropsy and histopathology. Mortality, malformations, and growth are endpoints assessed in FETAX.

A successfully validated FETAX could serve as a screening assay within a tiered scheme (e.g., a negative FETAX study would be followed by an *in vivo* mammalian assay, a positive FETAX study would not require further testing) to identify potential human teratogens and developmental toxicants. In this role, the assay has potential benefits with regard to reducing animal use and the cost and time associated with testing for developmental toxicants.

Test Method Protocol: A comprehensive guideline for conducting FETAX was published in 1991 under the auspices of the American Society for Testing and Materials (ASTM), as a "Standard Guide for Conducting the Frog Embryo Teratogenesis Assay—*Xenopus* (FETAX)," Annual Book of ASTM Standards, Designation E1439-91. In 1998, a revised ASTM FETAX Guideline (Designation E 1439-98) was produced. This guideline appears to be adequate to properly guide an investigator through the necessary test components and to ensure consistency in the testing methodology. One aspect of the protocol that may merit further investigation is the decision criterion used to identify a teratogenic response in FETAX. Several approaches have been suggested for improving the performance characteristics of FETAX compared to mammalian teratogenicity. One potentially significant improvement would be to base the $EC_{50}$ on characteristic malformations only, rather than on all malformations detected as is done currently. Characteristic malformations would be those that increase in frequency and possibly severity with increasing concentrations of the test substance.

Characterization of the Materials Tested: FETAX test data from 276 studies involving 137 substances, excluding environmental samples, were located, reviewed, extracted, and entered into the NICEATM FETAX database. The five most numerically prevalent chemical classes, in descending order, were nitrogen heterocyclic compounds (40 substances), amides and hydrazides (29 substances), organic (phenolic and carboxylic) acids (24 substances), alcohols (including glycols) (22 substances), and salts (20 substances). The five major product classes, in

descending order, were pharmaceuticals (45 substances), chemical synthesis components (17 substances), pesticides (13 substances), food additives (11 substances), and dyes (7 substances). In a number of cases, the same substance was placed in more than one chemical or product class.

Reference Data Used for Performance Assessment: Laboratory mammal (rat, mouse, and/or rabbit) reference data were located for 90 of the 137 substances and one environmental sample tested in FETAX. Human data (epidemiological and case-report information) were obtained for 31 of the 137 substances tested in FETAX and mammalian data were located for 30 of these. The quality of the data in terms of accuracy and whether the studies were conducted in compliance with national/international Good Laboratory Practice (GLP) guidelines was not determined.

Test Method Data and Results: The 1991 ASTM FETAX Guideline, with minor exceptions, was followed in the FETAX studies considered by NICEATM. All 137 substances in the FETAX database had been tested using without metabolic activation; 35 had also been tested with metabolic activation. Except for the most recent four of the five FETAX validation studies, blind coding was not used in any study to eliminate potential bias. Also, FETAX studies were not conducted in compliance with national or international GLP guidelines. The effect of these two issues on FETAX data quality is difficult to ascertain.

Test Method Performance Assessment: The performance characteristics (i.e., accuracy[1], sensitivity[2], specificity[3], positive predictivity[4], negative predictivity[5], false positive rate[6], and false negative rate[7]) of FETAX against rat, mice, and/or rabbit teratogenicity test results or human teratogenicity study results were determined by NICEATM. The decision criteria used in determining the performance characteristics of FETAX included single decision criteria (TI >1.5; TI >3.0; MCIG/$LC_{50}$ <0.30) and multiple decision criteria (TI >1.5 plus MCIG/$LC_{50}$ <0.30; TI >3.0 plus MCIG/$LC_{50}$ <0.30). When a multiple decision criterion was used, test substances were

[1] Accuracy: The proportion of correct outcomes of a method. Often used interchangeably with concordance.
[2] Sensitivity: The proportion of all positive chemicals that are correctly classified as positive in a test.
[3] Specificity: The proportion of all negative chemicals that are correctly classified as negative in a test.
[4] Positive Predictivity: The proportion of correct positive responses among materials testing positive.
[5] Negative Predictivity: The proportion of correct negative responses among materials testing negative.
[6] False Positive Rate: The proportion of all negative substances that are falsely identified as positive

classified as positive when both the TI value was greater than the decision point (1.5 or 3.0) and the MCIG/$LC_{50}$ ratio was less than 0.3; equivocal when one, but not both, criterion were positive; and negative when neither criterion was positive.

The performance characteristics of FETAX (with and/or without metabolic activation) was determined against all three laboratory mammal species (rat, mouse, and rabbit) combined or against each species alone. Using a single decision criterion, optimal performance for FETAX, with and without metabolic activation, compared against combined laboratory mammal data was based on a TI value greater than 1.5 (**Table A**). Using a multiple decision criterion did not enhance the performance characteristics of FETAX. Similar performance characteristics were obtained against rat, mouse, or rabbit, when considered individually.

**Table A. Performance Characteristics of FETAX**

| Performance Characteristics | FETAX, with and without metabolic activation, compared to Combined Laboratory Mammal (using TI >1.5) | FETAX, with and without metabolic activation, compared to Human (using MCIG/$LC_{50}$ <0.30) | Combined Laboratory Mammal compared to Human |
|---|---|---|---|
| Accuracy | 61% (55/90)* | 70% (19/27) | 63% (19/30) |
| Sensitivity | 82% (41/50) | 67% (8/12) | 71% (10/14) |
| Specificity | 35% (14/40) | 73% (11/15) | 56% (9/16) |
| Positive Predictivity | 61% (41/67) | 67% (8/12) | 59% (10/17) |
| Negative Predictivity | 61% (14/23) | 73% (11/15) | 69% (9/13) |
| False Positive Rate | 65% (26/40) | 27% (4/15) | 44% (7/16) |
| False Negative Rate | 18% (9/50) | 33% (4/12) | 29% (4/14) |

*Numbers in parenthesis indicate the number of accurate results/total number of substances compared.

[7] False Negative Rate: The proportion of all positive substances falsely identified as negative.

The performance of FETAX (with and/or without metabolic activation) was compared against human teratogenic data. Again, both single and multiple decision criteria were evaluated. Optimal performance was based on using a single decision criterion of an MCIG/$LC_{50}$ ratio less than 0.30. The resulting performance characteristics are presented in **Table A**. Using a multiple decision criterion did not significantly increase the performance characteristics of FETAX compared to human teratogenicity study results.

Maximal performance characteristics for laboratory mammal data compared to human teratogenicity results were obtained using rat, mouse, or combined laboratory mammal teratogenicity data, but not using rabbit data alone. The analysis was limited to substances tested in FETAX. The combined laboratory rat, mouse, and rabbit results are provided for comparative purposes in **Table A**.

The performance characteristics of FETAX, with and/or without metabolic activation, was determined for chemical and product classes that contained at least 15 substances with corresponding laboratory mammal or human teratogenicity results. Compared to laboratory mammal data, chemical and product classes evaluated included amides (15 comparisons), amides plus hydrazides (19 comparisons), amines (16 comparisons), amines plus nitrogen heterocyclic compounds (25 comparisons), nitrogen heterocyclic compounds (29 comparisons), phenolic and carboxylic acids (21 comparisons), and pharmaceuticals (40 comparisons). Compared to human study data, chemical and product classes evaluated included nitrogen heterocyclic compounds (17 comparisons) and pharmaceuticals (22 comparisons). Regardless of the single decision criterion used, performance characteristics were not appreciable different from those determined for the total database.

NICEATM evaluated the optimal TI value or MCIG/$LC_{50}$ ratio to use as a single decision criterion in FETAX for identifying teratogenic activity. Performance characteristics (accuracy, sensitivity, specificity) were determined against combined laboratory mammal (rat, mouse, and rabbit) or human teratogenicity results. Accuracy based on using either a TI value or an MCIG/$LC_{50}$ ratio as the single decision criterion value was never greater than ~60%, while a sensitivity of at least 85% was accompanied by a specificity of 40% or less. Differences in

performance characteristics between this analysis and the previous analysis reflect differences in the manner in which FETAX test results for the same substance from multiple studies were considered. In this analysis, the median TI value or MCIG/$LC_{50}$ ratio were used; in the previous analysis, a weight-of-evidence approach was used to classify results as positive or negative. The values obtained suggest that the use of FETAX as a screen, based on current decision criteria, is problematic.

The inclusion of a MAS in FETAX is considered essential for predicting developmental hazard in mammals. However, selection of the substances tested with a MAS do not appear to have been based on whether or not metabolic activation was thought to be required for teratogenic activity *in vitro*. Of the 35 substances tested with metabolic activation, only four are known to require metabolic activation to be reactive *in vitro*. Based on the limited database, additional studies to validate the role of metabolic activation in FETAX appear to be justified.

Several approaches have been suggested for modifying the decision criteria to increase the ability of FETAX to correctly identify developmental toxicants. These include:

- an evaluation of the $EC_{50}$ based on characteristic malformations (i.e., those increasing in incidence and severity with increasing test substance concentration) only,

- the calculation of a point estimate for the dose that inhibits growth by 50% rather than using an MCIG, and

- the use of 95% confidence intervals for statistically identifying TI values (and other point estimates) that are significantly different from the decision criteria value.

The effect of these approaches on the performance characteristics of FETAX has yet to be evaluated.

Test Method Reliability (Reproducibility/Repeatability): Five separate but related inter-laboratory FETAX validation studies in three phases were conducted. A total of 26 substances

were tested without metabolic activation and 14 substances with metabolic activation, with three to six different laboratories participating in each validation study (**Table B**). The Phase I Validation Study was classified as a training and protocol evaluation phase; the 1991 ASTM FETAX Guideline was followed. The subsequent four validation studies followed the same guideline with minor modifications (e.g., different preparation scheme for adding the test substance; 20 and not 25 embryos per dish when plastic rather than glass Petri dishes were used).

Validation was measured using the four different measurements obtained from FETAX—$LC_{50}$, $EC_{50}$, TI, and the MCIG. The investigators assessed reliability of each FETAX endpoint by calculating the coefficient of variation (CV) and conclusions about reliability were made from evaluating the range of CVs for each measure across laboratories (**Table B**). Additionally, the ASTM E691-92 (ASTM, 1992) guideline on a statistical approach for assessing intra- and inter-laboratory performance was used to evaluate test method reliability.

In the validation studies, there was excessive variability in the $LC_{50}$, $EC_{50}$, TI, and especially the MCIG within and across laboratories. A formal investigation into the factors contributing to this excessive variability has not been conducted. The resulting variation in these endpoints contributed to poor concordance among laboratories in regard to the classification of a test substance as a FETAX positive or negative, even when highly experienced laboratories were involved (**Table B**). A possible factor contributing to the variation in results may be that the types and severity of malformations are not currently included in the decision criteria used to classify substances as teratogenic or not teratogenic. A subsequent revision of the decision criteria emphasizing critical, or characteristic, malformations has been proposed.

Test Method Data Quality: Studies were conducted using routine laboratory practices, including standard record-keeping procedures. Studies were not conducted in accordance with Good Laboratory Practice (GLP) guidelines, nor were they generally conducted at facilities at which GLP studies are normally conducted. A quality assurance (QA) data audit of the FETAX Phase III.3 Validation Study indicated that data trails, study records, and results analysis procedures were not sufficient to support a standard GLP QA audit. An analysis of the accuracy of the data in the published report revealed the presence of occasional transcriptional errors; however, none

**Table B. Summary of FETAX Validation Studies**

| | Phase I (Bantle et al., 1994a) | Phase II (Bantle et al., 1994b) | Phase III.1 (Bantle et al., 1996) | Phase III.2 (Fort et al., 1998) | Phase III.3 (Bantle et al., 1999) |
|---|---|---|---|---|---|
| Number of Substances Tested | 3 | 4 | 6 | 2 | 12 |
| Number of Participating Laboratories | 7[a] | 7[a] | 7[a,b] | 7[a] | 3 |
| Tested Without MAS | Yes | Yes | Yes | Yes | Yes |
| Tested With MAS[c] | No | No | No | Yes | Yes |
| Coded Substances Used | No | Yes | Yes | Yes | Yes |
| Dose Selection Process | Common Doses | Common Doses | Individual Laboratory Selected | Individual Laboratory Selected | Individual Laboratory Selected |
| Overall CV mean and range (%), without MAS | 66.3 (20.5-201.5) | 24.4 (7.3-54.7) | 134.5 (21.7-991.6) | 26.0 (15.0-47.0) | 38.0 (9.5-87.2) |
| Overall CV mean and range (%), with MAS | N/A | N/A | N/A | 51.0 (18.0-131.0) | 51.1 (2.3-166.6) |
| Proportion of Study Results in Agreement (TI >1.5)[d] | 3 of 3 (100%) | 4 of 4 (100%) | 1 of 6 (17%) | 2 of 4 (50%) | 12 of 24 (50%) |
| Proportion of Study Results in Agreement (MCIG/$LC_{50}$ <0.30)[e] | 0 of 3 (0%) | 3 of 4 (75%) | 0 of 6 (0%) | 2 of 4 (50%) | 14 of 23 (61%) |

MAS = metabolic activation system.

[a] Six laboratories participated with one laboratory conducting each study twice using different technicians.

[b] Six studies instead of seven carried out evaluations for three of the six substances tested.

[c] Aroclor 1254-induced rat liver S9.

[d] Proportion of times that the participating laboratories agreed in classifying a FETAX study result as positive or negative, based on using a TI value greater than 1.5 as the single decision criterion.

[e] Proportion of times that the participating laboratories agreed in classifying a FETAX study result as positive or negative, based on using an MCIG/$LC_{50}$ of less than 0.30 as the single decision criterion.

of the discrepancies were considered to have significantly altered the reported general conclusions.

Other Scientific Reports and Reviews: No independent peer reviews of FETAX were located. Teratogenicity studies with *Xenopus* that did not follow the ASTM FETAX Guideline were located but excluded from consideration.

Animal Welfare Considerations: FETAX is proposed as a screen for human hazard identification (i.e., positive results only preclude the need for additional testing), and thus will not totally eliminate the use of mammals in teratogenicity and developmental toxicity testing. However, if accepted as a screen, use of this *in vitro* assay would reduce reliance on laboratory mammal tests, and thereby reduce the number of mammals used.

Other Considerations: Sufficient information on facilities and equipment for establishing FETAX is provided in the ASTM FETAX Guideline (1991, 1998). The three to six month estimated technical training time required for conducting the in-life portion of a FETAX study appears to be sufficient. However, based on concerns about differences in expertise in the identification of some of the more subtle malformations induced in *Xenopus* embryos, a more extensive training period may be required for the classification of malformations. The projected cost (<$25,000) and study duration (less than two months) for a Good Laboratory Practice (GLP) compliant FETAX study, with and without metabolic activation, appears to be reasonable. In comparison, a complete rat Prenatal Developmental Toxicity Study would cost about $120,000.

The potential impact of tetraploidy on the extrapolation of teratogenic changes in *X. laevis* to laboratory mammals and humans needs to be considered. Furthermore, the possible advantages of a diploid species of *Xenopus*, such as *X. tropicalis*, in FETAX, should be evaluated.

One recent development, which may greatly increase the utility of FETAX for identifying and prioritizing developmental hazards, is cDNA microarray technology. In this approach, developmental toxicity would be monitored at the level of the gene in terms of either up- or down-regulation. Given that exposures to different classes of developmental toxicants would be

expected to result in distinct patterns of altered gene expression, microarray technology could be utilized to categorize and classify these effects. In FETAX, treatment with a known developmental toxicant may provide a gene expression "signature" on a microarray, which represents the cellular response to these agents. When an unknown substance is tested, the microarray response could then be evaluated to see if one or more of these standard signatures is elicited. This approach might be used to elucidate an agent's mechanism of action, assess interactions between combinations of agents, or allow for a comparison between altered gene function in *Xenopus* with changes in analogous genes in mammalian systems. Currently, NIEHS is developing a custom "DNA chip" for *Xenopus* that is oriented toward the expression of genes involved in responses to toxic insults.

A number of *in vitro* systems have been proposed as alternatives or screens to *in vivo* mammalian developmental toxicity assays. A European Centre for the Validation of Alternative Methods (ECVAM)-sponsored validation of three *in vitro* assays considered suitable for the detection of substances posing a mammalian developmental hazard is in progress. The relative performance, cost-effectiveness, and flexibility of FETAX against other *in vitro* assays in identifying substances with mammalian developmental toxicity was not evaluated.

**FETAX For Ecotoxicological Hazard Assessment Using Water/Soil/Sediment Samples**

Rationale: Due to varying susceptibilities among animals, testing in multiple species is considered necessary to protect the environment. For each species, it is a combination of toxicants, water quality, and the organism itself that defines the hazard for a specific concentration of a toxicant within defined water quality conditions. Ecotoxicological standards are generally based on the susceptibility of the adult animal, which may not provide adequate protection for embryonic development and reproduction in many species. It is inherently impossible to evaluate developmental toxicity without exposing animals throughout development and assessing for adverse effects in multiple life stages. and for Early embryonic and juvenile stages are often the most susceptible periods for the toxic effects of many environmental contaminants. Embryonic development in amphibians is sensitive to water quality. Because of this, FETAX has been used in ecotoxicological studies to evaluate the potential developmental

hazard of contaminated surface waters, sediments, waste site soils, and industrial wastewater and to evaluate the efficacy of wastewater treatment procedures. In this context, the resulting data can be used to identify and prioritize sites with increased developmental toxicity risks.

Test Method Protocol: The 1991 and the revised and expanded 1998 FETAX Guideline published by ASTM is detailed, comprehensive, and well structured. Known limits of use for FETAX with water/soil/sediment samples were not described, except it was stated that the test method is incompatible with environmental samples that alter the pH, hardness, alkalinity, and conductivity of the FETAX Solution beyond the acceptable range. Testing of solids is generally limited by the water solubility of the constituents. The effects of other physico-chemical properties (e.g., nitrate levels) on *Xenopus* embryonic development need to be evaluated.

Characterization of Water/Soil/Sediment Samples Tested in FETAX: FETAX test data from ten publications involving 124 water/soil/sediment samples were located, reviewed, extracted, and entered into the NICEATM FETAX Environmental Sample Database.

Reference Data Used for an Assessment of FETAX Performance Characteristics: With one exception, laboratory mammal teratogenicity data for water/soil/sediment samples were not available, while relevant data for humans was nonexistent. Appropriate reference data for non-mammalian aquatic species was limited to a direct comparison in one sediment study and two-related soil extract studies between FETAX and *Pimephales promelas* (fathead minnow). Future ecotoxicological studies with FETAX should include tests on at least one reference species.

FETAX Test Method Data and Results: No attempt was made to obtain original data for any ecotoxicological study considered in this BRD. Generally, coded water/soil/samples were used for ease of identification and chain of custody. These studies were not conducted in compliance with GLP guidelines, nor were they generally conducted at facilities at which GLP-compliant studies are normally conducted. All 124 environmental samples in the NICEATM Environmental Sample Database had been tested using FETAX without metabolic activation; no environmental sample was tested also with metabolic activation.

FETAX has been used to evaluate the developmental toxicity of discharges from abandoned lead and zinc mines, contaminated ground and surface water samples collected near a closed municipal landfill, and direct discharges from industries and municipal wastewater treatment plants. This assay has also been used to assess the potential cause(s) of malformations and abnormalities observed in various species of frogs inhabiting bodies of water throughout the United States. FETAX has been used to assess the comparative hazard of soil samples from multiple waste sites contaminated with metals, PAHs, petroleum products, and organochlorine pesticides. The assay has also been used to test a series of five related fossil fuel mixtures as potential environmental pollutants.

Based on the studies evaluated, FETAX appears to be useful in ecotoxicological studies, and as a means for detecting and prioritizing sites with increased developmental hazard. Studies including other bioassays as part of a battery indicated that FETAX was sensitive enough to detect low levels of developmental abnormalities, but robust enough to be suitable for testing aqueous soil extracts. To increase the validity of the interpretation of such data, it may be useful to further evaluate the influence of the physico-chemical properties of environmental samples on the frequency of malformations in FETAX. Additionally, further research on the performance of the current FETAX protocol as an effective assay for assessing water and sediment quality and detecting changes that can have adverse effects on the ecosystem may provide further insight that could optimize ecotoxicological assessments. It would also be helpful to further evaluate how FETAX could best fit into a test battery for prioritizing of sites for further testing and remediation.

Performance Characteristics of FETAX with Water/Soil/Sediment Samples: Given the lack of sufficient reference data for comparison, the performance characteristics of FETAX, based on tests conducted using water/soil/sediment samples, could not be determined. However, there may be ecotoxicological testing applications where reference data for other species may not be needed or appropriate.

Test Method Reliability (Repeatability/Reproducibility): Due to the lack of appropriate inter-laboratory validation studies, an assessment of test method reliability with environmental

samples could not be conducted. One potential issue affecting data interpretation connected with water/soil/sediment samples is the lack of an exogenous MAS incorporated into the FETAX assay. An MAS would be useful where results are being used to predict effects on mammalian species. A FETAX validation study designed to evaluate test method reliability for ecotoxicological applications would be helpful. Such a study should include assessments by several laboratories, and should include the testing of both common samples and environmental samples collected independently. Studies focusing on data interpretation issues could also be helpful in further optimizing the assay. Potential issues to address include the decision criteria used for ranking samples in regard to developmental hazard, and the appropriateness of sample handling and processing techniques. ICCVAM Submission Guidelines should be followed in the design, conduct, and reporting of such studies.

Test Method Data Quality: Studies were not conducted in compliance with national or international GLP guidelines, nor were they generally conducted at facilities at which GLP studies are normally conducted. No data audits were conducted on studies testing environmental samples.

Other Scientific Reports and Reviews: No independent peer reviews of FETAX were located. Other data may exist that might be considered in an evaluation of the performance characteristics of FETAX for identifying developmental hazards in environmental samples.

Animal Welfare Considerations: Multiple species are generally used for ecotoxicological studies. Use of this *in vitro* assay could reduce reliance on tests involving adult organisms.

Other Considerations: Sufficient information on facilities and equipment for establishing FETAX is provided in the ASTM FETAX Guideline (1991, 1998). The estimated three to six month technical training time required for conducting the in-life portion of a FETAX study appears to be sufficient. However, based on concerns regarding the level of expertise needed for the proper identification of malformations induced in *Xenopus* embryos, more intensive training may be needed for this aspect of the assay. The projected cost (<$12,500) and study duration (<two

months) for a GLP compliant complete FETAX study, without metabolic activation, following the ASTM FETAX Guideline (1998), appears to be reasonable.

Other Applications for *Xenopus*: Other tests using *Xenopus* are being evaluated for their ability to identify substances or environmental samples that may disrupt endocrine function (the *Xenopus* Tail Resorption Assay, Vitellogenin Assay), for assessing reproductive toxicity, and for exploring limb mal-development, including possible mechanisms of action (*Xenopus* Limb Bud Assay). These developing test methods require appropriate validation.